Appendix A
US441 Piecewise Linear Regression

Introduction

A piecewise linear regression model was evaluated for the US 441 gage to determine if an improved fit could be obtained for flows near zero. This equation is of the form

When FtWhite_Q<knot1
    US441_Q = 0 + b3*Worthington_Q_lag3
When FtWhite_Q<knot2
    US441_Q = 0 + b1*(FtWhite_Q-knot1) + b3*Worthington_Q_lag3
When FtWhite_Q>=knot2
    US441_Q = 0+ b1*(FtWhite_Q-knot1) + b2*(FtWhite_Q-knot2) +
    b3*Worthington_Q_lag3

Where:

- US441_Q = daily flow at Santa Fe River at the US Hwy 441 Near High Springs gage (02321975)
- Worthington_Q_lag3 = daily flow at Santa Fe River at Worthington Springs gage (02321500), lagged three days
- FtWhite_Q = daily flow at Santa Fe River at the Fort White gage (02322500)
- Knot1 and knot2 are inflection points for flow at Fort White where the linear form slope changes
- b1, b2 and b3 are regression coefficients

Estimated model parameters were developed using the non-linear procedure in the statistical software SPSS (Version 16.0) and are included in Table 1 and Attachment A. Selected exceedance flow estimates for the model period of record (POR) (1992-May 1, 2019) for the US 441 gage data and the piecewise model is included in Table 2.

The inflection values (knots) are only related to Fort White flow – flow at Worthington Springs is weighted by a constant coefficient (b3). For flows less than knot1 (559.4 cfs) the slope associated with flow at Fort White is zero. Between a flow of 559.4 cfs and knot2 (1479.9 cfs) the slope is 0.682, which means for every cfs increase in flow at Fort White, flow at US 441 increases by 0.682 cfs, plus the influence of flow at Worthington Springs. When flows at Fort White are greater than 1479.9, the slope decreases slight from 0.682 by 0.142. For example, if flow at Fort White is less than 559.4 cfs (knot1) and flow at Worthington Springs is zero, then flow at US441 is estimated to be zero. The small decline in slope at Fort White flows greater than about 1480 cfs likely is due some relative loss of water (i.e., decreased gain) between US 441 and Fort White under high hydraulic head in the river. The piecewise linear regression model is in good agreement with the data across a range of flows from zero to between 5% and 1% exceeded where it underestimates about 10 % (Table 2).

Confidence intervals

Confidence intervals (i.e. of the mean response flow at US441) were estimated around fitted values of the piecewise linear regression model using a standard error of the fit approach assuming normal residuals, and a bootstrapped approach to assess the influence of a non-normal distribution of residuals. The bootstrapped approach (1) randomly sampled model residuals with replacement (i.e. duplicates allowed), (2) added the resampled residuals to observed flow values at US441, and (3) refit the

piecewise regression model 10,000 times. The fitted values (10,000 bootstrap replicates × 9,572 flow values) were used to estimate confidence intervals since they represent a random distribution of residuals that do not follow the original (non-normal) distribution of residuals. The bootstrapped confidence intervals were estimated as the lower 2.5% and upper 97.5% of residuals calculated for each flow value.

The resulting 95% confidence intervals not assuming a normal distribution of residuals (i.e. bootstrapped confidence intervals) were compared to the 95% confidence intervals assuming a normal distribution of residuals (i.e. non-bootstrapped confidence intervals). The widths of bootstrapped and non-bootstrapped confidence intervals are similar (Figure 1), with average widths of 12 cfs (bootstrapped) and 15 cfs (non-bootstrapped) and widths of approximately 9 cfs (bootstrapped) and 11 cfs (non-bootstrapped) around the median fitted value.

The similarity in confidence intervals follows the central limit theorem in that a random sample drawn from a sufficiently large sample, such as 9,572 flow values, will generate results that conform to a normal distribution. This applies here as the original large sample of flow values and the bootstrapped, random sample generated residuals evenly centered around fitted values, with minor deviations around that central point (like a normal distribution around the mean).

Table 1. Parameter estimates for the US441 gage using a 2-knot piecewise linear regression

Parameter	Parameter Estimate	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
knot1	559.4	8.9	541.9	576.8
b1	0.682	0.013	0.656	0.708
knot2	1479.9	71.299	1340.2	1619.7
b2	-0.142	0.014	-0.169	-0.115
b3	0.425	0.006	0.412	0.437

Table 2. Estimated flow exceedances calculated using different models for US441 compared to gaged flow for water years 1993-2019 (through May 1)

Model[2]/Exceedance	1	5	10	25	50	75	90	95	99
US441 gaged flow data	3701	1648	1212	696	356	133	33	0	0
Piecewise 2 knots (Model 2[1])	3347	1693	1191	695	352	148	33	7	0
[1]Model 2 is naming convention used in SPSS.									

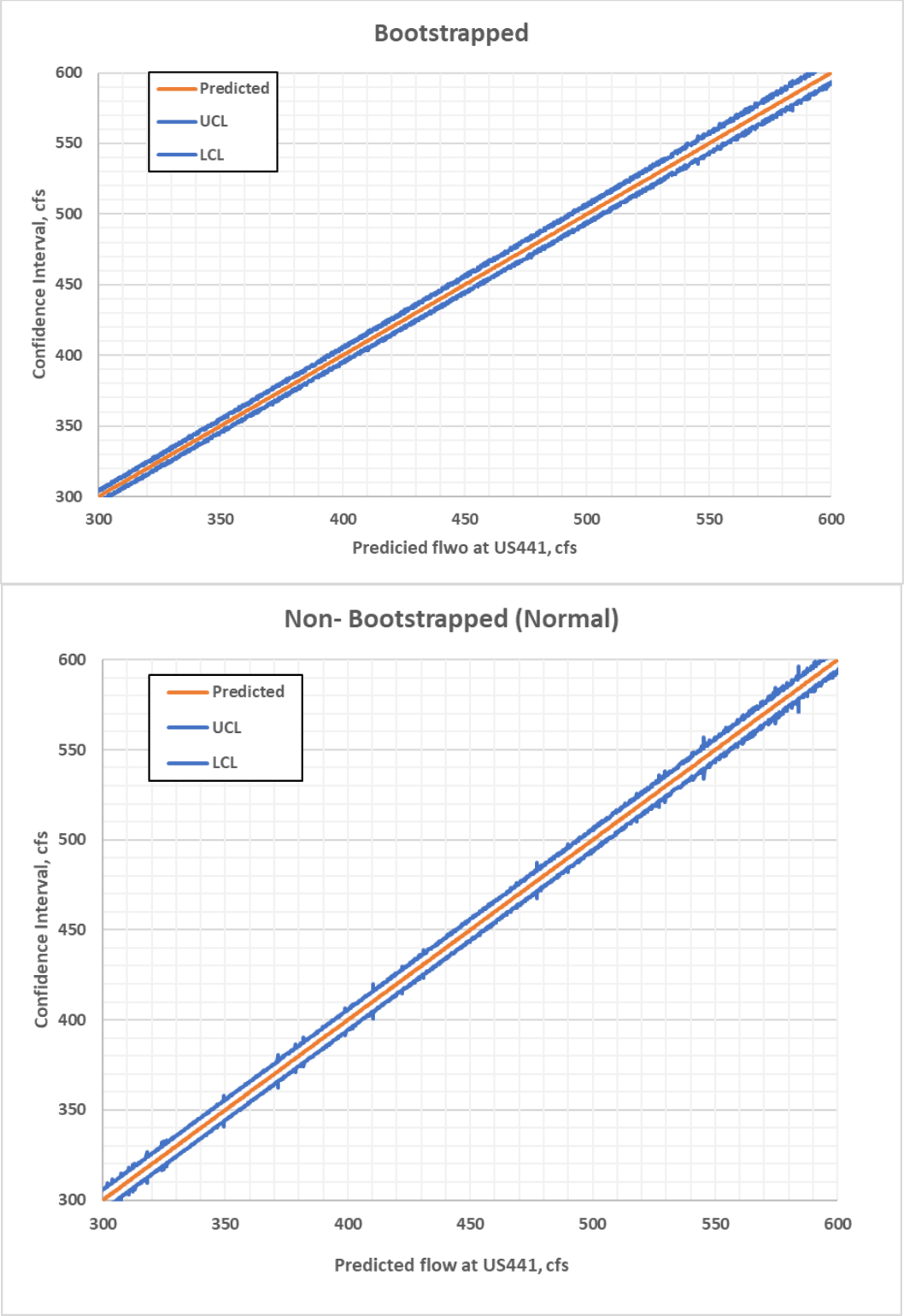


Figure 1. Bootstrapped and non-bootstrapped (normal distribution) confidence intervals around fitted values of the piecewise linear regression model output for the US441 gage.

Attachment A

Nonlinear Regression Analysis (Model 2 Knots)

```
* Non-Linear Regression.
MODEL PROGRAM knot1=500 b1=.6 knot2=900 b2=0 b3=.4.
COMPUTE  PRED_=0*(FtWhite_Q<knot1)+ b1*(FtWhite_Q-knot1) * (FtWhite_Q>=knot1)+
b2*(FtWhite_Q-knot2) * (FtWhite_Q>=knot2) + b3*Worthi
   ngton_Q_lag3.
NLR US441_Q
  /OUTFILE='C:\Users\kww\AppData\Local\Temp\spss193488\SPSSFNLR.TMP'
  /PRED PRED
  /SAVE PRED RESID
  /CRITERIA SSCONVERGENCE 1E-8 PCON 1E-8.
```

Iteration History[b]

Iteration Number[a]	Residual Sum of Squares	Parameter				
		knot1	b1	knot2	b2	b3
1.0	7.029E8	500.000	.600	900.000	.000	.400
1.1	6.976E8	587.092	.797	900.000	-.236	.421
2.0	6.976E8	587.092	.797	900.000	-.236	.421
2.1	6.986E8	564.470	.668	1108.146	-.113	.422
2.2	6.986E8	564.470	.668	1108.146	-.113	.422
2.3	6.947E8	573.374	.712	1080.532	-.157	.422
3.0	6.947E8	573.374	.712	1080.532	-.157	.422
3.1	6.911E8	561.275	.681	1332.380	-.134	.424
4.0	6.911E8	561.275	.681	1332.380	-.134	.424
4.1	6.901E8	559.790	.682	1471.048	-.140	.425
5.0	6.901E8	559.790	.682	1471.048	-.140	.425
5.1	6.901E8	559.303	.682	1482.477	-.142	.425
6.0	6.901E8	559.303	.682	1482.477	-.142	.425
6.1	6.901E8	560.131	.684	1465.029	-.143	.425
6.2	6.901E8	559.466	.683	1472.498	-.142	.425
6.3	6.901E8	559.273	.682	1477.732	-.142	.425
7.0	6.901E8	559.273	.682	1477.732	-.142	.425
7.1	6.901E8	559.303	.682	1482.386	-.142	.425
7.2	6.901E8	559.316	.682	1478.952	-.142	.425
8.0	6.901E8	559.316	.682	1478.952	-.142	.425
8.1	6.901E8	559.353	.682	1481.605	-.142	.425
8.2	6.901E8	559.334	.682	1479.390	-.142	.425
9.0	6.901E8	559.334	.682	1479.390	-.142	.425
9.1	6.901E8	559.367	.682	1480.283	-.142	.425
9.2	6.901E8	559.346	.682	1479.668	-.142	.425
10.0	6.901E8	559.346	.682	1479.668	-.142	.425
10.1	6.901E8	559.367	.682	1480.232	-.142	.425
10.2	6.901E8	559.351	.682	1479.795	-.142	.425
11.0	6.901E8	559.351	.682	1479.795	-.142	.425
11.1	6.901E8	559.361	.682	1480.051	-.142	.425

11.2	6.901E8	559.356	.682	1479.918	-.142	.425
12.0	6.901E8	559.356	.682	1479.918	-.142	.425
12.1	6.901E8	559.365	.682	1480.165	-.142	.425
12.2	6.901E8	559.357	.682	1479.949	-.142	.425

Derivatives are calculated numerically.

a. Major iteration number is displayed to the left of the decimal, and minor iteration number is to the right of the decimal.

b. Run stopped after 34 model evaluations and 12 derivative evaluations because the relative reduction between successive residual sums of squares is at most SSCON = 1.00E-008.

Parameter Estimates

Parameter	Estimate	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
knot1	559.357	8.900	541.911	576.804
b1	.682	.013	.656	.708
knot2	1479.949	71.299	1340.188	1619.710
b2	-.142	.014	-.169	-.115
b3	.425	.006	.412	.437

Correlations of Parameter Estimates

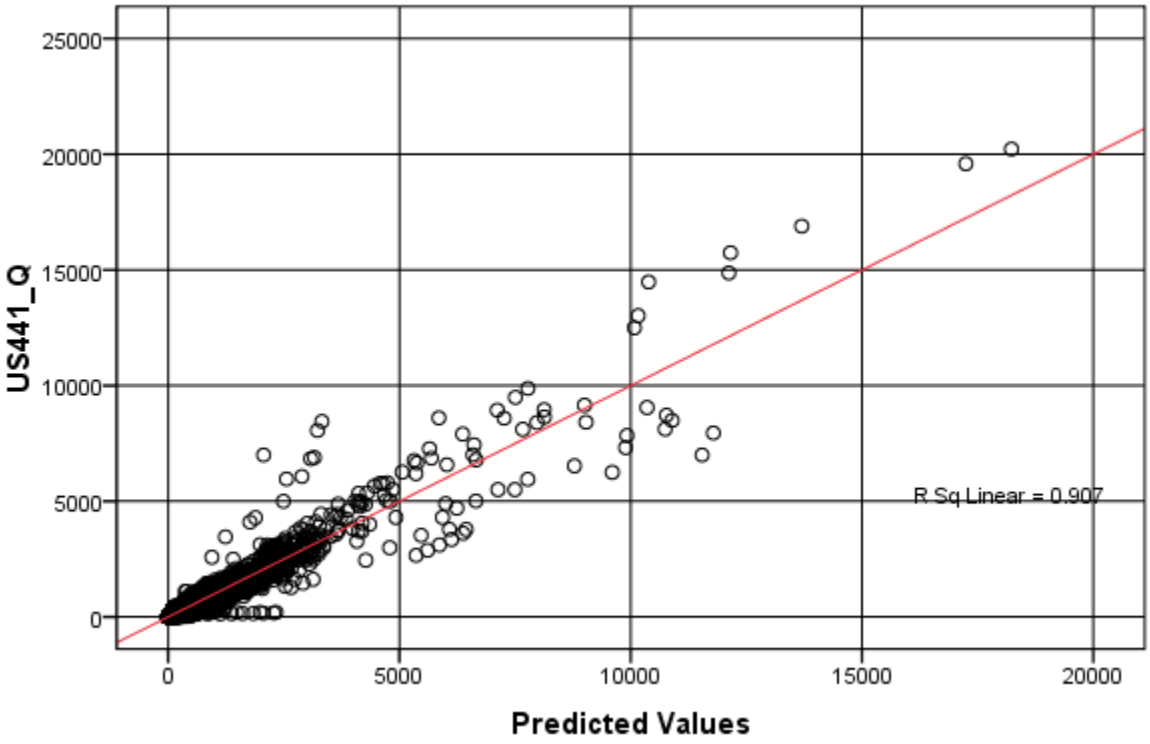
	knot1	b1	knot2	b2	b3
knot1	1.000	.837	-.393	-.788	-.029
b1	.837	1.000	-.642	-.885	-.189
knot2	-.393	-.642	1.000	.546	-.071
b2	-.788	-.885	.546	1.000	-.192
b3	-.029	-.189	-.071	-.192	1.000

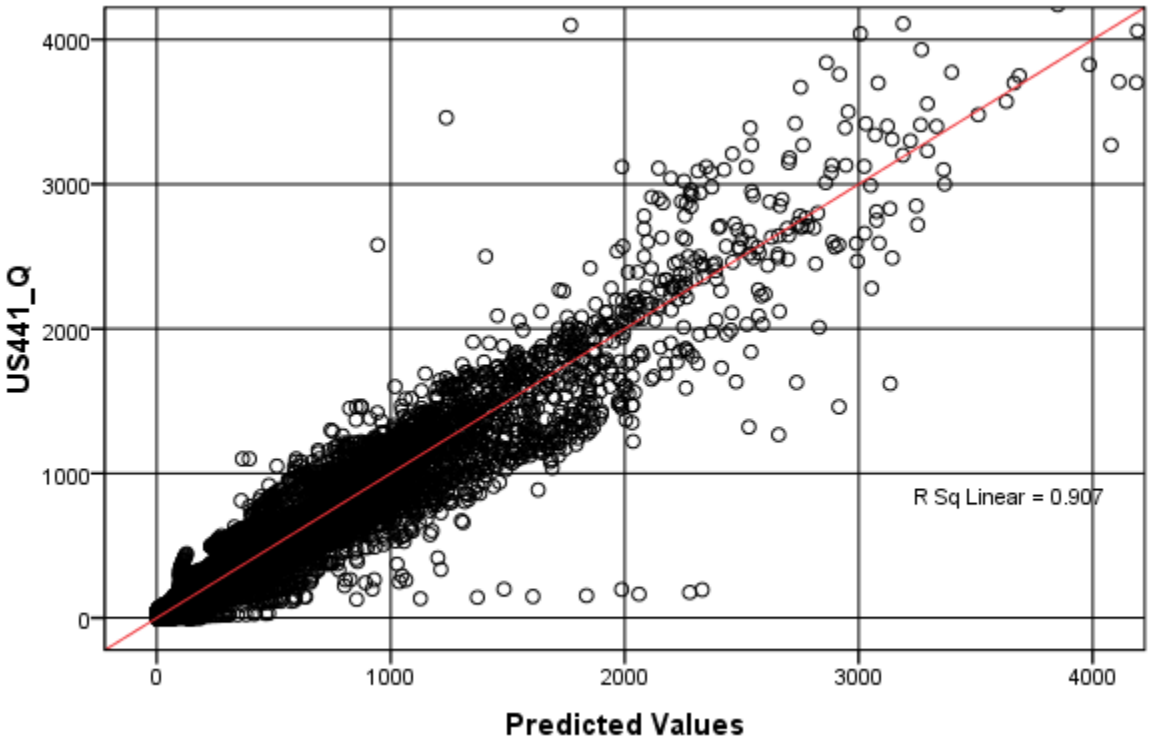
ANOVA[a]

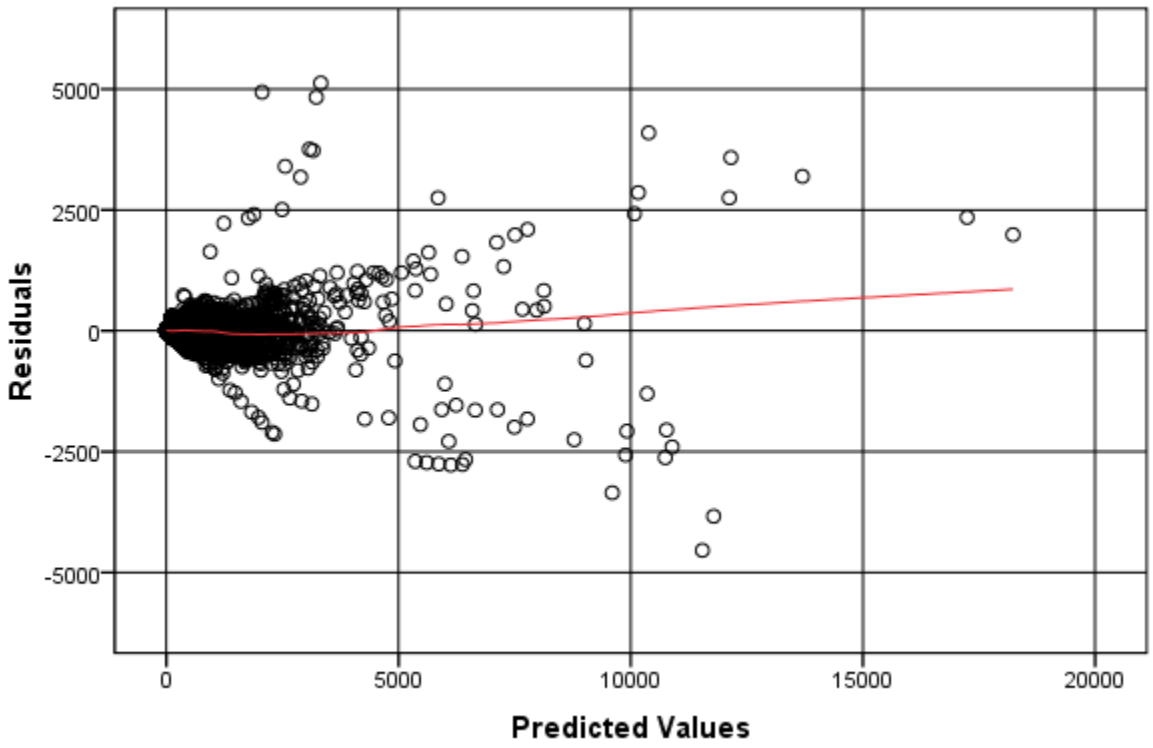
Source	Sum of Squares	df	Mean Squares
Regression	9.865E9	5	1.973E9
Residual	6.901E8	9567	72136.485
Uncorrected Total	1.056E10	9572	
Corrected Total	7.454E9	9571	

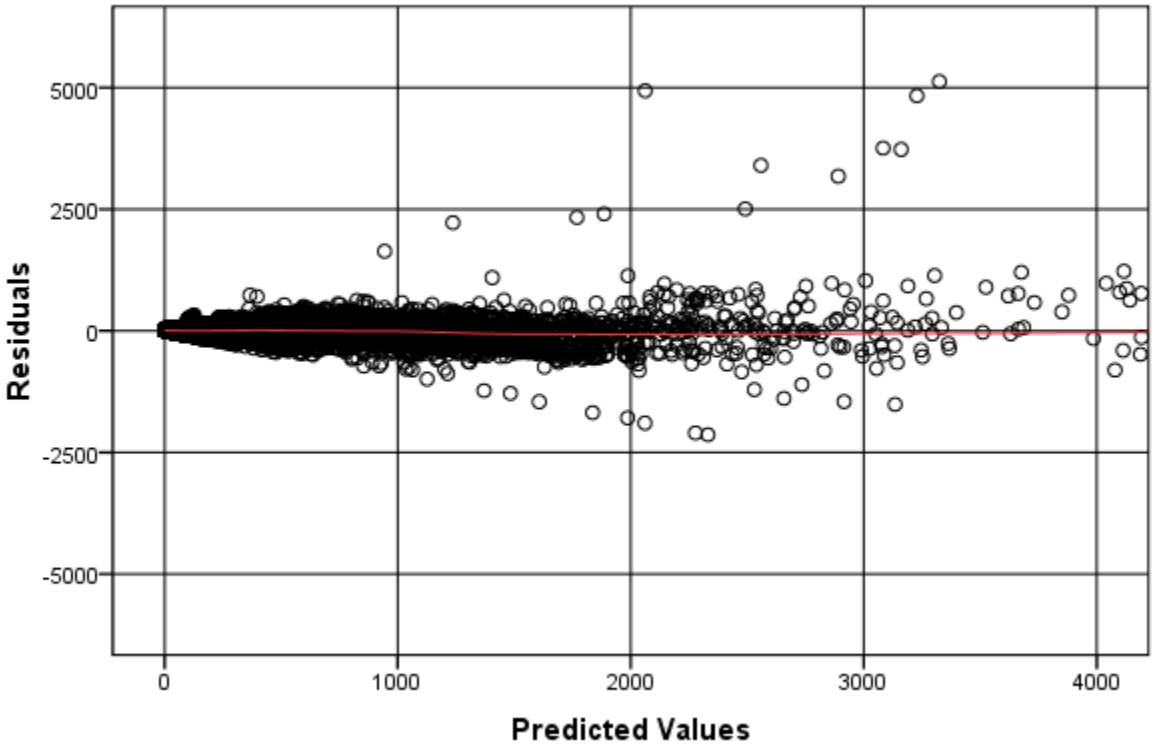
Dependent variable: US441_Q

a. R squared = 1 - (Residual Sum of Squares) / (Corrected Sum of Squares) = .907.


US441_Q
25000
20000
15000
10000
5000
0
0
5000
10000
15000
20000
Predicted Values
R Sq Linear = 0.907


US441_Q
4000
3000
2000
1000
0
0
1000
2000
3000
4000
Predicted Values
R Sq Linear = 0.907


Residuals
5000
2500
0
-2500
-5000
0
5000
10000
15000
20000
Predicted Values


Residuals
5000
2500
0
-2500
-5000
0
1000
2000
3000
4000
Predicted Values